DOCUMENT RESUME

ED 455 255 TM 033 057

AUTHOR	Rabinowitz, Stanley; Ananda, Sri
TITLE	Balancing Local Assessment with Statewide Testing: Building a Program That Meets Student Needs. Knowledge Brief.
INSTITUTION	WestEd, San Francisco, CA.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	2001-00-00
NOTE	9p.
CONTRACT	RJ96006901
AVAILABLE FROM	WestEd, 730 Harrison Street, San Francisco, CA 94107-1242. Tel: 877-493-7833 (Toll Free); Web site: http://www.WestEd.org.
PUB TYPE	Opinion Papers (120)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Elementary Secondary Education; Models; Program Implementation; Psychometrics; *School Districts; *State Programs; *Test Construction; Test Use; *Testing Programs

ABSTRACT

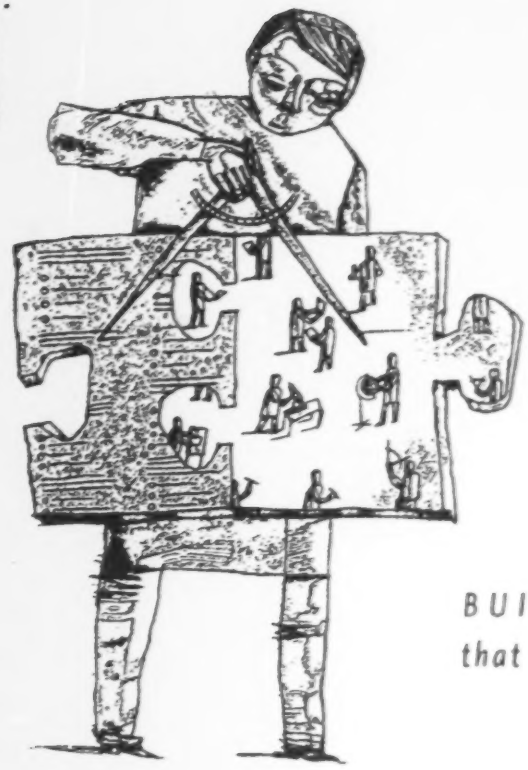
This brief makes the case for thoughtful district- or school-driven assessment systems that complement and go beyond what statewide testing systems are able to accomplish. It describes important attributes of model local assessment programs and presents the necessary steps for building a local assessment program that will elicit information that is of value to teachers, students, and parents and is rarely available from state assessment programs. Heavily influencing the development of nationwide assessments are issues related to the technical adequacy of assessments and their efficiency. These issues are more easily managed at the local level, where assessments are rarely used for graduate or system accountability purposes. Despite fewer constraints related to technical adequacy or efficiency, many local officials have been tempted to develop systems that essentially duplicate their state's assessment program. Efficient and effective local assessments will complement, rather than duplicate, statewide efforts. Such assessments should be linked to state and local content standards, provide information values at the local level, and support teaching and learning. These steps will ensure the development of efficient and effective local assessment systems: (1) identify and prioritize needs and goals; (2) meet with state assessment officials; (3) identify resources; (4) convene development teams; (5) provide necessary professional development; (6) pilot tasks and reports; (7) revise tasks based on pilot results; and (8) implement and monitor. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

knowledge brief

ED 455 255

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.



© WILLIAMSON 2000

BALANCING Local Assessment with Statewide Testing:

BUILDING A PROGRAM that meets Student Needs

Students across America are being tested at unprecedented rates, due in large part to a proliferation of state developed and administered assessments. Forty-eight of the nation's 50 states have adopted some form of statewide assessment program, collectively spending hundreds of millions of dollars annually on increasingly complex systems. In the past, states tended to test only the so-called basics of language arts and mathematics. Today, many students are being tested in additional academic areas, such as science and social studies, as well as in nontraditional content, such as workplace readiness. Complicating the picture are the high stakes associated with much of the assessment: For students, test results may affect grade promotion or graduation; for schools or districts, they can trigger accountability-related rewards and sanctions. In this sea of statewide, high-stakes assessment, it's logical to wonder if and how *local* assessment fits in.

Written by Stanley Rabinowitz and Sri Ananda

TM033057

WestEd
Improving education through research, development, and service

A local program consists of a formal set of assessment approaches and tools developed or selected by school districts or, in some cases, individual schools to meet their own needs. This is distinct from assessments developed by an individual classroom teacher for his or her own purposes, such as end-of-unit tests or the Friday quiz. Developing and implementing a local system requires extensive expertise, time, and money — in other words, a lot of effort. Given the degree and type of statewide testing, are local programs still worth that effort? The answer is a resounding yes, but only if key criteria are met.

This brief makes the case for thoughtful district- or school-driven assessment systems that complement, and go beyond, what statewide testing systems are able to accomplish. It

BEST COPY AVAILABLE

describes important attributes of model local assessment programs. Finally, it presents the necessary steps for building a local assessment program that will elicit information that is of value specifically to teachers, students, and parents and that is rarely available from state assessment programs.*

State Vs. Local Assessment: A Role for Each

As statewide assessment programs focus increasingly on high-stakes student and school accountability concerns, they must rely increasingly on narrower and more conservative assessment methods, primarily multiple-choice tests. The strong suit of these instruments is their ability, in a valid, reliable, and efficient manner, to reveal patterns of relative strengths and weaknesses across large groups of students. Such information can serve as an early warning system, pointing to content areas, schools, student groups, and even individual students warranting greater attention. What such statewide tests generally do not yield is specific-enough data to use in targeting instruction for individual students. This leaves a clear and essential role for local assessment: developing diagnostic information about what students do well, where they are having difficulty, and how the instructional program might be adjusted to address their specific needs.

Local assessment programs have greater potential for generating this kind of complex information largely because they are not bound by the same constraints as state-level programs. As a result, they can more realistically incorporate innovative assessment methods, such as portfolios and performance events, which are able to generate more specific information about the strengths and weaknesses of individual students.

STATE-LEVEL LIMITATIONS

Heavily influencing the development of statewide assessments are two overlapping issues: the technical adequacy of assessments and their efficiency.

Technical Adequacy. In a high-stakes testing environment, assessment instruments must demonstrate sufficient technical quality to support accountability decisions (e.g., student retention, promotion, graduation, teacher awards, school sanctions); otherwise, the assessment agency risks litigation. With regard to high school graduation testing and other student accountability measures, for example, legal rulings have set a very high technical bar, requiring strong evidence of reliability, validity, access, and lack of bias. While many years of research and development have gone into innovative assessment approaches, such as the use of projects, portfolios, or running records, those approaches cannot easily match the technical quality of traditional testing methodologies. The technical adequacy of multiple-choice testing, or multiple-choice testing coupled with some short constructed-response items, remains better understood, easier to demonstrate, and therefore more practical for state-level assessment.

> Innovative assessment approaches, such as the use of projects, cannot easily match the technical quality of traditional testing methodologies.

Among the many examples of how state testing has become increasingly conservative to support high-stakes policies is Kentucky's decision to drop performance

* Rabinowitz and Ananda (2000) explain the reasons behind the growth in statewide assessment programs and describe what a model state program might look like in *A Model Student Assessment System to Support School Accountability*, a paper presented at the Council of Chief State School Officers' Annual Conference on Large-Scale Assessment, Snowbird Village, Utah.

events and the mathematics portfolio from its assessment and accountability system. To further increase the reliability of the system, it has added the norm-referenced Comprehensive Tests of Basic Skills. Another example is California's interim decision to base its accountability system on Stanford 9 scores because more innovative approaches required more time to develop and implement.

Efficiency. Feeling pressured by public frustration about large numbers of poorly performing schools, under-prepared college freshmen, and ill-prepared entry-level workers, many state policymakers — governors, legislators, state boards of education — seek changes that will yield visible improvements quickly. More and more, they are looking to school and student accountability systems, and they want these systems in place *now.* This sense of urgency tends to rule out performance-based assessments, which take longer to develop than multiple-choice tests and are generally more costly to implement.

For example, multiple-choice tests can be machine scored in very little time at nominal cost. By contrast, statewide scoring of student essays, projects, and portfolios takes far more time and can cost millions of dollars because it involves human scorers who must be trained, with their work calibrated and monitored. Moreover, even if assessment development time was not an issue, some states would still hesitate to use such methods because their implementation would be seen by many as encroaching on precious instructional time. Add to this the cost of teacher professional development in how to implement performance assessment and it's easy to understand why states are choosing to rely instead on traditional assessment methods, even if they only measure global performance of students and school systems.

Despite fewer constraints, many local officials have been tempted to develop systems that essentially duplicate their state's assessment program.

LOCAL-LEVEL OPPORTUNITY

While the technical adequacy and efficiency of assessment are also issues at the local level, they are more easily managed. Rarely are locally developed assessments used for graduation or system accountability purposes. This lowers the technical requirements for assessment instruments, and a broader range of evidence can justify their use. For example, the somewhat lower reliability of a performance task may be counter-balanced by its higher content validity and consequential validity. From an efficiency perspective, with student graduation on the line, the technical bar for a state test could require that student essays be read twice, each time by a different scorer. At the local level, when the principal purpose of the assessment is diagnosis, essays might be read only once by the students' teacher, thereby saving money and ensuring assessment results in a more timely fashion. Also, performance tasks are best implemented and managed at the classroom level. Locally developed systems are able to involve a larger percentage of affected classroom teachers at all phases of the development and implementation process, increasing their buy-in. Finally, locally developed tasks ensure the greatest degree of match between what is valued at the local level and what is assessed.

Yet, despite fewer constraints related to technical adequacy or efficiency, many local officials have been tempted to develop systems that essentially duplicate their state's assessment program, using identical tools and focusing on the same content, just at different grades. The perceived logic is this: Because state assessment programs typically measure student achievement in only selected grades — for example, 4, 8, and 11 — using the same instrument to measure achievement in all other grades would allow schools to develop complete trend lines for all students. In theory,

this would measure annual individual growth. Yet for the majority of students, little grade-to-grade variation occurs in performance on standardized tests, whether the tests are norm-referenced or criterion-referenced; such tests are not designed to show reliable individual growth in these relatively small increments. So testing all students for this purpose would waste time and money. A more sensible reason for using comparable assessment instruments to test in off grades would be to predict later performance in the accountability grades. Schools could use test results to identify likely student performance deficits, then make programmatic or instructional changes to address these performance gaps before performance actually "counted" for accountability purposes. However, even this could be more efficiently accomplished by testing only at certain grades (as opposed to testing *all* grades) for prediction purposes. It could also be accomplished by using a valid local predictive alternative, such as teacher observation.

> Local assessments should provide detailed diagnostic information for each student because state tests provide only basic global information at the student level.

Preferable still is a more targeted approach to local assessment overall, one that reflects good assessment practice and is consistent with requirements for federal and state compensatory education programs. The model system would allow schools to concentrate limited resources on in-depth assessment and analysis of those students and content areas most in need of attention.

Attributes of a Model Local Assessment Program

As implied above, effective and efficient local assessment programs will complement, rather than duplicate, statewide efforts. Moreover, they are responsive to local constituencies, including students, parents, teachers, administrators, and the community at large. In building or revising a local assessment program, local policymakers and teachers, working together, should ensure that the system has the following attributes:

Linked to State and Local Content Standards. Ideally, state content standards reflect knowledge and skills that are appropriate for all students and measurable on a statewide assessment. But local communities might value additional content or skills that would not meet those criteria. They could, for example, have their own standards reflecting local values and economic needs. Thus, local curriculum and, therefore, assessments might reflect different content or a different emphasis than that embodied in the state standards and assessment. In its new state content standards, Nevada actually designates which standards are appropriately assessed at the state level — in the state graduation test and other required statewide tests — and which are best addressed locally because they must be assessed with more innovative methods.

Provide Information Valued at the Local Level. Local assessments should provide detailed diagnostic information for each student because state tests, either by virtue of the design chosen or due to inherent methodological constraints, provide only basic global information at the student level. A typical state mathematics test, for example, provides a reasonable measure of whether a student is generally strong or weak in mathematics. It can also provide a moderately reliable assessment of a student's relative strengths at the sub-score level (e.g., computation vs. problem solving, algebra vs. geometry). But it can provide little useful data on how to address performance weaknesses. By contrast, a well-designed local assessment can supply diagnostic information. Districts or schools might choose one of two approaches to filling the diagnostic void, depending on the extent of the achievement gap

BEST COPY AVAILABLE

Considering Performance Assessments? Four Key Questions

Local officials must determine when performance-based assessments are the best tool for accomplishing important instructional and accountability goals. In many cases, they will decide that sufficient information can be obtained from multiple-choice assessment (or from multiple-choice plus constructed-response tasks). The following questions can help in making this determination:

- **Is there a need for the evidence a performance assessment can provide?** Performance tasks are expensive and time-consuming to develop, implement, score, and report. Local programs should develop or adapt assessment tasks only for those content areas in which students are known to be performing poorly and/or on students who have been performing below standard. It may also make sense to use them to assess areas in which local decision-makers are simply not satisfied with the data yielded by other types of tests.

- **Can you ensure assessment results that are timely and user-friendly?** A common complaint about the use of performance assessment in statewide programs is the length of time from administration to reporting. This is due mainly to the hand-scoring requirement. The same problem can plague local efforts. Results must be available when needed for important decisions: designing a student's education plan, placing students in an appropriate program, or determining if an instructional program should be continued or revised. Because, as noted earlier, lower stakes allows greater scoring flexibility locally, the time from administration to reporting could be shortened. For instance, a classroom teacher could score the tasks for her own students and there might be no need for a second person to score the same tasks. Equally important, results must be provided in a user-friendly format for students, teachers, and parents. Because performance events yield more complex, unwieldy, and unfamiliar information than that obtained from multiple-choice tests, care must be taken in the design and interpretation of reports for intended audiences.

- **Is this assessment affordable?** Many great assessment ideas are poorly implemented because planners have underestimated the effort and resources required to implement them. Teachers tend to underestimate the amount of time necessary for students to complete complex tasks, while administrators tend to underestimate the degree of support required for teachers and students to be successful (e.g., professional development on how to teach more complex content). For this reason, experienced consultants should be used as needed in the design phase and at key checkpoints throughout the process, for example, at the point of developing a scoring report. Also, it is often better to begin using performance assessments in one grade and one content area rather than jumping headfirst into all subjects across grades. This more targeted approach requires setting clear priorities. Although this may ruffle some feathers among those who feel their students or content areas are being left behind, the consequences of trying to move ahead in a less focused manner can be a legacy of failure — and skepticism about any assessment innovation.

- **Is performance assessment "worth it"?** Even if they have answered the first three questions positively, local staff should always ask themselves whether there is a more efficient method of getting the information they need about student learning. How much better must a performance-based approach be to justify its use over a traditional multiple-choice counterpart? This question can only be answered through analyzing both needs and available resources. Costs need to be considered not just in fiscal terms, but in terms of lost or gained opportunities (e.g., what other things would we *not* be able to do if we developed these assessments? What would be the cost of failure?).

When the answer to all of these questions is yes, performance assessment can play a key role in a local program. And when performance assessments are combined with locally developed or adapted multiple-choice assessments — or when their results are considered in conjunction with helpful data from state-level tests — the result is a coherent, local assessment system that provides the ideal balance to state-level testing.

among their students and the extent of available resources. They can administer more detailed assessments to all students with the intent of building a tailored education plan for each, or they can focus attention on students identified by the state test as achieving below standard, then concentrate resources and diagnostic attention on this smaller pool. This latter approach is more efficient because it takes advantage of reliable information from the state test to identify the most pressing needs, allowing schools to concentrate their more limited resources on this targeted student population.

Support Teaching and Learning. Large-scale, system-monitoring assessments at the state level don't tend to promote thoughtful classroom practices. In fact, they often result in a narrowing of instruction as teachers focus on raising test scores. And because the format of most state tests is largely multiple choice, in their attempt to prepare students for the test, teachers may give less instructional attention to certain higher-order skills (e.g., conceptual understanding). Free from some of the constraints of a state-level program, a local assessment program has greater potential to promote more effective teaching and learning. It can do so by using performance-based assessment tools, such as projects, demonstrations, journals, students' self-evaluations, and/or portfolios, to support greater development of students' metacognitive abilities (e.g., problem solving, critical reasoning, application of knowledge in real-world contexts).

The Development of Local Assessment Systems

Several comprehensive guides are available to help local educators develop and implement assessments designed for specific goals, student populations, and content areas (Assessment Laboratory Network, 2000; O'Neill & Stansbury, 2000; Stiggins, 1999). What follows is a brief overview of the steps that ensure the most efficient and effective implementation of local assessments. These steps are necessary irrespective of the instruments chosen, and of whether the school or district decides to develop its own assessments or use or adapt existing tests. Note that, as a general rule, the more innovative the program, the more time and effort are necessary for successful implementation. Schools should plan on 12–18 months to develop and pilot potential assessments before they can be implemented. In many cases, it might be best to stagger the development process across several school years, rather than attempt to simultaneously implement all components of a local system.

1. **Identify and Prioritize Needs and Goals.** The needs that the local assessment system is expected to address and its expected outcomes should be identified as early as possible. Only then can staff decide what combination of assessment instruments is appropriate. In making that decision, it's important to consider the concept of ***value added***: Is the assessment being proposed worth the time and effort of students and teachers? Is there another less costly way of getting the information sought? How would this assessment work contribute to raising the achievement of *all* students, particularly those most at risk? Having considered these questions, policymakers then need to meet with and gather the support of constituencies within and beyond the school walls (e.g., teachers, parents, business leaders). Most important at this point is developing a process by which decisions will be made and resources found and allocated. Lead staff must be identified, trained, and empowered.

2. **Meet With State Assessment Officials.** Before investing in a new assessment system, local staff should meet with their counterparts at the state level who deal with both assessment policy and technical issues. This commonly overlooked step can yield several advantages. First, a thorough understanding of the state program, including its future directions, can ensure that the local program is complementary, not duplicative. Next, state officials might be able to identify other local agencies that have embarked on similar development activities. Finally, the state may be able to allocate technical staff and other resources to assist in the local effort.

3 **Identify Resources.** Local development takes time and money, including the costs of shifting staff from other activities. Budgets need to be developed. Outside sources of funding (e.g., businesses, four dations) may be required. External technical consultants might be needed. Some existing testing instruments may be available that can be adopted or adapted, resulting in substantial savings. In some instances, policymakers must decide whether ongoing, repetitive tasks, such as scoring and reporting, should be an internal function or contracted out. An excellent resource may be other schools or districts with similar goals and plans; if so, a consortium can be formed to pool talent and resources and to create other significant savings and efficiencies. A well-developed plan (see step 1) is essential for a realistic estimate of the human and fiscal resources needed and for their appropriate allocation.

4 **Convene Development Teams.** While existing instruments may be available, chances are that some additional development will be necessary. In almost every case, the use of development teams, provided with a proper charge and training, will improve the final product, as compared to the results of an individual working alone. The use of consultants familiar with the test-development process can be invaluable at this point. Teams should consist of teachers, administrators, and when appropriate, parents and other community members. This makeup will result in both more valid tasks and broader support for their implementation.

5 **Provide Necessary Professional Development.** The professional development needs of teachers expected to implement the new system must be considered as early as step 1. A complex system that no one can implement is doomed. Professional development activities fall into four general categories: (a) the philosophy and goals of the local assessment system; (b) how to teach consistent with that philosophy; (c) how to administer the actual assessments, including scoring; and (d) how to interpret results, for teachers, students, parents, and administrators. Training might need to be repeated over time to reach newly hired staff and to refresh the knowledge of existing teachers and administrators.

6 **Pilot Tasks and Reports.** All new tasks must undergo a pilot-test process. This "dress-rehearsal" will ensure that tasks work as expected and teachers, students, and support staff are ready for the new expectations. Piloting can also identify specific content that teachers might have thought they were teaching well, but for which assessment scores show otherwise. This information can then lead to changes in curriculum and/or instruction.

7 **Revise Tasks Based on Pilot Results.** Invariably, glitches occur. Some tasks may take longer to administer than expected or not work at all. Others may not be equally suitable for all segments of the student population (e.g., low-performing students). Revision time, often substantial, must be built into the implementation schedule.

8 **Implement and Monitor.** Over time, the new system should run more smoothly. Indicators of success should be developed and regularly monitored throughout the development and implementation process.

The above process can be complex. But careful adherence can result in a local assessment program that complements its state counterpart in goals, focus, and approach. Properly developed and implemented, a local system can yield truly valuable information about student learning — information that can guide instruction and program development, ultimately resulting in higher achievement. And when it complements the state system, a local assessment program can yield data that support reform efforts without overburdening students, teachers, and the education system in which they operate.

For a complete listing of WestEd's products, visit our Web site at

WestEd.org

REFERENCES

Assessment Laboratory Network. (2000). *Making assessment work for everyone: How to build on student strength*. San Francisco: WestEd.

O'Neill, K., & Stansbury, K. (2000). *Developing a standards-based assessment system*. San Francisco: WestEd.

Stiggins, R. J. (1999). *Student-centered classroom assessment*. New York: Merrill.

WestEd is a nonprofit research, development, and service agency that works with education and other communities to promote excellence, achieve equity, and improve learning for children, youth, and adults. Drawing on the best knowledge from research and practice, our agency collaborates with practitioners, policymakers, and others, addressing critical education and other human service issues. WestEd, with initiatives throughout the United States and abroad, is one of the nation's designated Regional Educational Laboratories — originally created by Congress in 1966 — serving the states of Arizona, California, Nevada, and Utah. With headquarters in San Francisco, WestEd has offices across the United States.

For more information about WestEd, visit our Web site at WestEd.org; call 415.565.3000 or, toll-free, (1-877) 4WestEd; or write: WestEd, 730 Harrison Street, San Francisco, CA 94107-1242.

This report was produced in whole or in part with funds from the Office of Educational Research and Improvement, U.S. Department of Education, under contract #RJ96006901. Its contents do not necessarily reflect the views or policies of the Department of Education.

© 2001 WestEd. All rights reserved. Permission to reproduce, with WestEd copyright notice included, is hereby granted.

WestEd®

730 Harrison Street
San Francisco
California 94107-1242

Address service requested

Non-Profit
U.S. Postage
PAID
Los Alamitos, CA
90720
Permit No. 87

BEST COPY AVAILABLE